Synthesizing Research on the Generation and Maintenance of Population Diversity

Emily Dolson and Charles Ofria

BEACON Center for Evolution in Action, Michigan State University, 48824
Department of Computer Science and Engineering, Michigan State University, 48824
Program in Ecology, Evolutionary Biology, and Behavior, Michigan State University, 48824
dolsonem@msu.edu

Abstract

The concept of diversity has different definitions, usages, and nuances when looking from one field to another. Evolutionary biologists are primarily interested in the population dynamics that produce diversity, ecologists want to understand the maintenance and community-level effects of diversity, and evolutionary computation researchers want to exploit diversity to produce better and more varied solutions to real-world problems. In artificial life, we are particularly interested in understanding diversity as a critical component of natural systems in order to produce artificial ones that exhibit comparable open-ended dynamics. Here we begin to develop a framework to unite these views on diversity, with a goal of facilitating the transfer of ideas among these fields and formulating a consistent vocabulary.

Introduction

Population diversity is of critical importance to many fields adjacent to artificial life. Because evolution requires meaningful variation for selection to act on, diversity is essential for harnessing the constructive power of an evolving system, such as is used to solve challenging evolutionary computation problems (Burke et al., 2004); these techniques may also be useful to improve crop breeding, animal husbandry, or directed evolution of proteins (Bull and Wichman, 2001). Similarly, the generation of new diversity is a topic of substantial interest in evolutionary biology, as it enables continued adaptation to new niches. Meanwhile, ecologists are fascinated with questions about existing diversity: How so much diversity is able to stably coexist (Chase and Leibold, 2003; Chesson, 2000)? What are the implications of having a diverse ecosystem (Tilman et al., 2014)?

Artificial life lies at the intersection of these fields and provides a useful perspective from which to synthesize ideas, measurements, and applications of diversity. The field fosters systems that can evolve and sustain ecologies, and provides a common language that facilitates translating results back and forth across disciplinary boundaries.

All of these different questions about diversity are interrelated. The mechanisms that lead to diversity being generated influence the mechanisms that are able to maintain it. Different mechanisms for diversity maintenance and generation lead to qualitatively different types of diversity that are more or less conducive to evolving solutions to any particular problem (Mouret and Doncieux, 2009). While there are many different kinds of diversity that are each studied independently, many of them have strong correlations with each other. Thus, although there is a great deal of research focused on "diversity", this research is far more fragmented than it should be. For example, we know a lot about how evolutionary dynamics produce *de novo* diversity and a lot about how diversity affects ecosystem functioning, but the lack of a common framework prevents us from directly inferring how those same evolutionary dynamics affect ecosystem functioning.

Here, we introduce such a framework consisting of two components: an approach to classifying the vast array of diversity concepts, and an approach to thinking about how these concepts interact with each other. Our classifications indicate the type of target being measured (e.g. genes, phenotypes, or species), the specific metrics used for the diversity measurement (e.g., Shannon diversity or phylogenetic distance), and how the collection of targets for these measurements is chosen (by region, by attribute, by ancestry, etc.). In examining how these concepts of diversity interact, we pay special attention to when in evolutionary history they are at play; for example, some induce more mutational variation in a population allowing for the production of new diversity, while others require pre-existing interactions among species in order to manifest.

Measuring diversity

There are three components to any measurement of diversity: the **type** of categorization, the specific **measurement** being taken, and the scale of the **collection** within which diversity is being measured. Any valid diversity metric consists of a combination of type, measurement, and collection.

Type What are the units that we want to measure the diversity of? These can be, for instance, genotypes, phenotypes, behavioral types, or functional types. Importantly, all of these types have much more precise definitions in biology

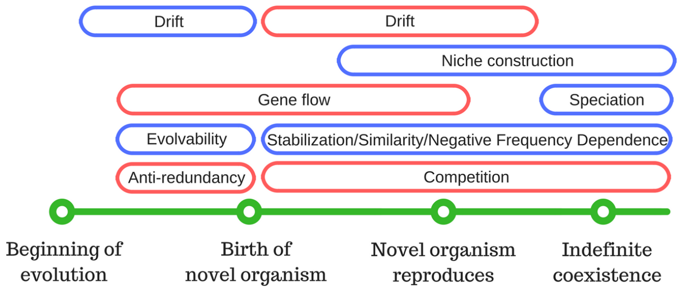

Figure 1: The different factors affecting the generation and maintenance of diversity over evolutionary time. The green timeline indicates the relative timing of important reference events. Ovals indicate different mechanisms that promote (blue) or inhibit (red) diversity generation/maintenance. Placement of the ovals relative to the timeline indicate when these processes operate.

than in evolutionary computation (with artificial life falling somewhere in between). The only hard and fast rule in computational systems is that the phenotype is determined by the genotype. Behavioral and functional types are generally a subset of what biologists consider the phenotype.

Collection What group are we calculating the diversity of? Is it a single taxonomic unit, like a species? Is it a collection of organisms living in a common spatial location? Is it a set of communities? Each of these collections is subject to different pressures. For example, gene flow is, typically, considered only within a single species.

Measurement Once we have a collection of units of a certain type, what calculation do we want to perform? In biology, the most common measurement is richness - the number of unique values of the type. In computational systems it is Shannon entropy, which combines evenness and richness. Similarity-based measurements, such as Levenshtein (edit) distance, are less common in both fields, but have the potential to be powerful descriptors; if a given evolutionary operator produces many very similar genotypes, for example, that would have important differences from producing more extreme variants. The more popular similarity metric in biology is phylogenetic similarity, which indicates how closely related two types are. Oddly, phylogenetic metrics are not frequently used in computational systems, despite the relative ease of tracking them.

Factors affecting diversity over time

Traditionally, factors that impact biodiversity have been split into those that help to initially generate it and those that help to maintain it once it exists. In reality, however, this dichotomy is an oversimplification. How early in the life of a novel taxon must a mechanism act to count as supporting its generation? Does diversity generation only include factors that maneuver the population into a part of the fitness landscape that is conducive to diversification? Or does it also include factors that prevent novel individuals from being rapidly eliminated before becoming established? We argue that it is more productive to place factors along a timeline (see Figure 1) and note their relative contributions.

This timeline perspective facilitates some interesting observations. For example, drift promotes the generation of novel mutations by allowing the population to explore more of the fitness landscape. However, from the perspective of a novel mutation that has just appeared, drift is a factor that is likely to eliminate this newly-generated diversity; the population of individuals with this mutation is necessarily small, and so stochastic effects have an out-sized influence on it.

Conclusions

Going forward, we expect these approaches to thinking about diversity across fields to benefit a wide variety of diversity researchers. To that end, we are in the process of assembling a large-scale review synthesizing diverse research and perspectives on diversity. We expect that this process will clarify the interrelationships among concepts of diversity in different fields, and enable us to determine the most pressing open questions.

References

Bull, J. J. and Wichman, H. A. (2001). Applied Evolution. *Annual Review of Ecology and Systematics*, 32(1):183–217.

Burke, E. K., Gustafson, S., and Kendall, G. (2004). Diversity in genetic programming: An analysis of measures and correlation with fitness. *Evolutionary Computation, IEEE Transactions on*, 8(1):47–62.

Chase, J. M. and Leibold, M. A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press. Google-Books-ID: Ssmcl_ubQUQC.

Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, 31:343–366.

Mouret, J.-B. and Doncieux, S. (2009). Using Behavioral Exploration Objectives to Solve Deceptive Problems in Neuro-evolution. In *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, GECCO '09, pages 627–634, New York, NY, USA. ACM.

Tilman, D., Isbell, F., and Cowles, J. M. (2014). Biodiversity and Ecosystem Functioning. *Annual Review of Ecology, Evolution, and Systematics*, 45(1):471–493.